Blind Second-Order Source Extraction of Instantaneous Noisy Mixtures

Wei Liu, Danilo P. Mandic, and Andrzej Cichocki

Abstract—The problem of blind source extraction (BSE) for noisy measurements is addressed in the domain of second-order statistics using the linear predictor method. By extending the results from the noise-free case, two methods for the noisy case are proposed, whereby, for rigor, the effect of noise is removed from the cost function. The so introduced algorithms are based, respectively, on the minimization of the normalized mean square prediction error (MSPE), and the minimization of MPSE. The analysis of the derived BSE algorithms is supported by simulations.

Index Terms—Additive noise, blind source extraction (BSE), online learning, second-order statistics.

I. INTRODUCTION

BLIND source extraction (BSE) is a special class of blind source separation (BSS) methods, [1]–[3], which, instead of recovering all the source signals simultaneously, extracts only one or a subset of the sources at a time. As pointed out in [3], BSE has an obvious advantage of saving computation when we are not interested in all of the source signals. By BSE, we can also extract the sources in a prescribed order according to their statistical and other properties and choose different algorithms at different stages. Algorithms specifically designed for BSE can be roughly divided into two categories: those based on higher order statistics (HOS) [4]–[7] and those based on second-order statistics (SOS) [8]–[12], which assume that the sources are not correlated with each other and every source has a different temporal structure. The SOS approach is mainly based on a linear predictor and an adaptive algorithm was proposed in [8], whereas a batch algorithm was derived for a one-step predictor in [9]. In [12], a critical study of this approach showed analytically the role of prewhitening in this structure and based on the analysis, a new efficient algorithm was derived by minimizing the normalized mean square prediction error (MSPE).

Fig. 1 shows this structure, where the extracted signal $y[n]$ and the instantaneous output error $e[n]$ of the linear predictor with a length P are given by

$$y[n] = \mathbf{w}^T \mathbf{x}[n]$$
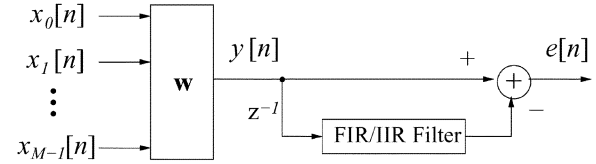$$e[n] = y[n] - \mathbf{b}^T \mathbf{y}[n] \qquad (1)$$

Fig. 1. BSE structure.

where $\mathbf{w}$ is the demixing vector, $\mathbf{x}[n]$ is the vector of the observed mixtures at time instant n, and

$$\mathbf{b} = [b_1 \quad b_2 \quad \cdots \quad b_P]^T$$
$$\mathbf{y}[n] = [y[n-1] \quad y[n-2] \quad \cdots \quad y[n-P]]^T$$
$$\mathbf{x}[n] = [x_0[n] \quad x_1[n] \quad \cdots \quad x_{M-1}[n]]^T. \qquad (2)$$

In general, $\mathbf{x}[n]$ is given by

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] + \mathbf{v}[n] \qquad (3)$$

where $\mathbf{A}$ is the $M \times L$ mixing matrix, and $\mathbf{s}[n]$ is the $L \times 1$ source signal vector given by

$$\mathbf{s}[n] = [s_0[n] \quad s_1[n] \quad \cdots \quad s_{L-1}[n]]^T$$
$$[\mathbf{A}]_{m,l} = a_{m,l}, m = 0, \ldots, M-1, \qquad l = 0, \ldots, L-1 \qquad (4)$$

whereas $\mathbf{v}[n]$ is the additive noise vector.

In the previously proposed linear predictor methods, the noise term was not taken into account in the derivation of the algorithms, however, noise is present in real world measurements. It has been proven that by minimizing the normalized MSPE $J_0(\mathbf{w}) = (E\{e^2[n]\})/(E\{y^2[n]\})$ [12], the sources can be extracted successfully. An alternative method is to minimize the MSPE $J_1(\mathbf{w}) = E\{e^2[n]\}$ directly, but this requires preprocessing in the form of prewhitening and a normalization of the demixing vector $\mathbf{w}$ during the adaptive process [3]. In this brief, we consider the case with additive noise and propose a new cost function, based upon which a novel class of BSE algorithms is derived.

The brief is organized as follows. In Section II, based on an analysis of the prediction error, we provide a new cost function which accounts for the presence of noise, after which we derive the proposed class of BSE algorithms. Simulation results are given in Section III and conclusions drawn in Section IV.

II. BSE WITH A LINEAR PREDICTOR: PROPOSED ALGORITHMS

A. Proposed Cost Function

The BSE structure using a linear predictor is shown in Fig. 1. For this approach to work, we need to impose some assumptions

on the statistics of the source signals, such as uncorrelatedness and different temporal structures, which are given by

$$
\begin{aligned}
\mathbf{R}_{ss}[0] &= E\{\mathbf{s}[n]\mathbf{s}^T[n]\} \\
&= \mathrm{diag}\{\rho_0[0], \rho_1[0], \ldots, \rho_{L-1}[0]\}
\end{aligned} \tag{5}
$$

with $\rho_m[0] = E\{s_m[n] \cdot s_m[n]\}, m = 0, 1, \ldots, L-1$, where $E\{\cdot\}$ denotes the statistical expectation operator and

$$
\begin{aligned}
\mathbf{R}_{ss}[\Delta n] &= E\{\mathbf{s}[n]\mathbf{s}^T[n - \Delta n]\} \\
&= \mathrm{diag}\{\rho_0[\Delta n], \rho_1[\Delta n], \ldots, \rho_{L-1}[\Delta n]\}
\end{aligned} \tag{6}
$$

with $\rho_l[\Delta n] \neq 0$ for some nonzero delay Δn.

Moreover, we assume the noise $\mathbf{v}[n]$ is uncorrelated with the source signals and has a correlation matrix given by

$$
\begin{aligned}
\mathbf{R}_{vv}[\Delta n] &= E\{\mathbf{v}[n]\mathbf{v}^T[n - \Delta n]\} \\
&= \begin{cases} \mathbf{0}, & \text{for } \Delta n \neq 0 \\ \mathbf{R}_v, & \text{for } \Delta n = 0. \end{cases}
\end{aligned} \tag{7}
$$

Normally, we further assume that $\mathbf{R}_v = \sigma_v^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix and σ_v^2 is the variance of the noise.

The linear predictor in the BSE structure can be either an finite-impulse response (FIR) or infinite-impulse response (IIR) filter, adaptive or with fixed weights [3]. Without loss of generality, we shall assume a P-tap FIR filter with coefficient vector $\mathbf{b}$. The MSPE of the structure from Fig. 1 can be expressed as [3]

$$
\begin{aligned}
E\{e^2[n]\} =\ & E\{y^2[n]\} - 2E\{y[n]\mathbf{b}^T\mathbf{y}[n]\} \\
& + E\{\mathbf{b}^T\mathbf{y}[n]\mathbf{y}^T[n]\mathbf{b}\} \\
=\ & \mathbf{w}^T\mathbf{R}_{xx}[0]\mathbf{w} - 2\sum_{p=1}^{P} b_p \mathbf{w}^T\mathbf{R}_{xx}[p]\mathbf{w} \\
& + \sum_{p,q=1}^{P} b_p b_q \mathbf{w}^T\mathbf{R}_{xx}[q-p]\mathbf{w}
\end{aligned} \tag{8}
$$

where $\mathbf{R}_{xx}[\Delta n] = E\{\mathbf{x}[n]\mathbf{x}[n - \Delta n]^T\}$ is the correlation matrix of the observed mixtures. After accounting for the effect of noise, which is uncorrelated with the source signals, from (3), we have

$$
\begin{aligned}
\mathbf{R}_{xx}[\Delta n] &= \mathbf{A}E\{\mathbf{s}[n]\mathbf{s}^T[n - \Delta n]\}\mathbf{A}^T \\
& \quad + E\{\mathbf{v}[n]\mathbf{v}^T[n - \Delta n]\} \\
&= \mathbf{A}\mathbf{R}_{ss}[\Delta n]\mathbf{A}^T + \mathbf{R}_{vv}[\Delta n].
\end{aligned} \tag{9}
$$

Thus, the MSPE $E\{e^2[n]\}$ can be divided into two parts

$$
E\{e^2[n]\} = e_s^2 + e_v^2. \tag{10}
$$

The term e_s^2 is attributed to the source signals and given by

$$
\begin{aligned}
e_s^2 =\ & \mathbf{w}^T\mathbf{A}\mathbf{R}_{ss}[0]\mathbf{A}^T\mathbf{w} - 2\sum_{p=1}^{P} b_p \mathbf{w}^T\mathbf{A}\mathbf{R}_{ss}[p]\mathbf{A}^T\mathbf{w} \\
& + \sum_{p,q=1}^{P} b_p b_q \mathbf{w}^T\mathbf{A}\mathbf{R}_{ss}[q-p]\mathbf{A}^T\mathbf{w} \\
=\ & \mathbf{w}^T\mathbf{A}\hat{\mathbf{R}}_{ss}\mathbf{A}^T\mathbf{w}
\end{aligned} \tag{11}
$$

where $\hat{\mathbf{R}}_{ss}$ is a diagonal matrix given by

$$
\hat{\mathbf{R}}_{ss} = \mathbf{R}_{ss}[0] - 2\sum_{p=1}^{P} b_p \mathbf{R}_{ss}[p] + \sum_{p,q=1}^{P} b_p b_q \mathbf{R}_{ss}[q-p]. \tag{12}
$$

The diagonal elements of the matrix $\hat{\mathbf{R}}_{ss}$ are the MSPEs associated with the corresponding source signals.

The term e_v^2 from (10) is caused by the noise, with

$$
e_v^2 = \mathbf{w}^T\mathbf{R}_v\mathbf{w} + \sum_{p=1}^{P} b_p^2 \mathbf{w}^T\mathbf{R}_v\mathbf{w} = \sum_{p=0}^{P} b_p^2 \mathbf{w}^T\mathbf{R}_v\mathbf{w} \tag{13}
$$

where b_0 is defined as $b_0 = 1$.

If there is no noise present, we can minimize the normalized MSPE $J_0(\mathbf{w}) = (E\{e^2[n]\})/(E\{y^2[n]\})$, or the MSPE $J_1(\mathbf{w}) = E\{e^2[n]\}$ if we can constrain the output power of the demixing vector $E\{y^2[n]\}$ to be unity. However, in the presence of noise, we cannot simply follow the same route to minimize J_0 or J_1. We need to remove the effect of noise in the cost function in such a way that the resulting cost function is the same as in the noise-free case.

Note, in the presence of noise

$$
\begin{aligned}
E\{y^2[n]\} &= \mathbf{w}^T\mathbf{R}_{xx}[0]\mathbf{w} \\
&= \mathbf{w}^T\mathbf{A}\mathbf{R}_{ss}[0]\mathbf{A}^T\mathbf{w} + \mathbf{w}^T\mathbf{R}_v\mathbf{w}.
\end{aligned} \tag{14}
$$

We therefore construct a new cost function J_2

$$
J_2(\mathbf{w}) = \frac{E\{e^2[n]\} - e_v^2}{E\{y^2[n]\} - \mathbf{w}^T\mathbf{R}_v\mathbf{w}}. \tag{15}
$$

Based on the above analysis, we next have

$$
J_2(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{A}\hat{\mathbf{R}}_{ss}\mathbf{A}^T\mathbf{w}}{\mathbf{w}^T\mathbf{A}\mathbf{R}_{ss}[0]\mathbf{A}^T\mathbf{w}}. \tag{16}
$$

Let $\mathbf{g} = \mathbf{A}^T\mathbf{w}$ denote the global demixing vector. Then (16) becomes

$$
J_2(\mathbf{w}) = \frac{\mathbf{g}^T\hat{\mathbf{R}}_{ss}\mathbf{g}}{\mathbf{g}^T\mathbf{R}_{ss}[0]\mathbf{g}}. \tag{17}
$$

Without loss of generality, we shall assume $\mathbf{R}_{ss}[0] = \mathbf{I}$, as the differences in the diagonal elements of $\mathbf{R}_{ss}[0]$ can always be absorbed into the mixing matrix $\mathbf{A}$. This way, the diagonal elements of $\hat{\mathbf{R}}_{ss}$ become the normalized MSPEs and they are assumed to be different from each other, otherwise the scheme will not be able to extract all the sources successfully. This yields

$$
J_2(\mathbf{w}) = \frac{\mathbf{g}^T\hat{\mathbf{R}}_{ss}\mathbf{g}}{\mathbf{g}^T\mathbf{g}}. \tag{18}
$$

Let us define a new vector $\hat{\mathbf{g}} = (\mathbf{g})/(\sqrt{\mathbf{g}^T\mathbf{g}})$, for which $\hat{\mathbf{g}}^T\hat{\mathbf{g}} = 1$. Then, the cost function (18) becomes

$$
J_2(\mathbf{w}) = \hat{\mathbf{g}}^T\hat{\mathbf{R}}_{ss}\hat{\mathbf{g}}. \tag{19}
$$

Consider now the optimization problem formulated as

$$\min_{\hat{\mathbf{g}}} \hat{\mathbf{g}}^T \hat{\mathbf{R}}_{ss} \hat{\mathbf{g}} \quad \text{subject to} \quad \hat{\mathbf{g}}^T \hat{\mathbf{g}} = 1. \tag{20}$$

The solution to this problem is a vector $\hat{\mathbf{g}}_{\text{opt}}$ with only one nonzero element, and strictly equal to unity at the position corresponding to the smallest diagonal element of the matrix $\hat{\mathbf{R}}_{ss}$ [3]. As $\hat{\mathbf{g}} = (\mathbf{g})/(\sqrt{\mathbf{g}^T \mathbf{g}})$, the corresponding global demixing vector $\mathbf{g}_{\text{opt}}$ will be the same as $\hat{\mathbf{g}}_{\text{opt}}$ except that the nonzero element in $\mathbf{g}_{\text{opt}}$ is an arbitrary constant c. Since we are minimizing $J(\mathbf{w})$ with respect to $\mathbf{w}$, instead of $\hat{\mathbf{g}}$, we need to prove that there exists a $\mathbf{w}_{\text{opt}}$ which results in $\hat{\mathbf{g}}_{\text{opt}}$.

In fact, from $\mathbf{g} = \mathbf{A}^T \mathbf{w}$, when $\mathbf{A}$ is of full rank and the number of mixtures M is larger or equal to the number of sources L, $\mathbf{w}_{\text{opt}}$ can be obtained using the pseudo-inverse of $\mathbf{A}^T$, as

$$\mathbf{w}_{\text{opt}} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{g}_{\text{opt}}. \tag{21}$$

As the possible minimum value of $J(\mathbf{w})$ is reached only when $\mathbf{g} = \mathbf{g}_{\text{opt}}$, as long as there exists such a $\mathbf{w} = \mathbf{w}_{\text{opt}}$ so that $\mathbf{g} = \mathbf{g}_{\text{opt}}$, we can state that when we minimize $J(\mathbf{w})$ with respect to $\mathbf{w}$, the result will be a successful extraction of the source signal with a minimum normalized MSPE.

The key to the success of this method is, we must have the knowledge of the the correlation matrix $\mathbf{R}_v$, which, in general, in real-world situations, is not realistic. However, as pointed out earlier, in practice, it is reasonable to assume $\mathbf{R}_v = \sigma_v^2 \mathbf{I}$. In this case, we have

$$\mathbf{R}_{xx}[0] = \mathbf{A}\mathbf{R}_{ss}[0]\mathbf{A}^T + \sigma_v^2 \mathbf{I}. \tag{22}$$

When $M > L$, that is, the number of mixtures is larger than the number of sources, we can use a subspace method to estimate σ_v^2, which in this case represents the smallest eigenvalue of $\mathbf{R}_{xx}[0]$ [13]. Alternatively, we can use an adaptive principal component analysis method to estimate σ_v^2 online [2]. In fact, if we simply minmize $E\{y^2\}$ subject to the constraint $\mathbf{w}^T \mathbf{w} = 1$, the optimum value of $E\{y^2\}$ will be σ_v^2.

B. Derived Algorithms

It has been proved that we can extract the sources successfully simply by minimizing the function $J_2(\mathbf{w})$. To simplify the derivation of the adaptive algorithms, we will perform a normalization of $\mathbf{w}$ after each update, given by

$$\mathbf{w}[n+1] \leftarrow \mathbf{w}[n+1]/\sqrt{\mathbf{w}^T[n+1]\mathbf{w}[n+1]} \tag{23}$$

which also helps avoid the critical case where the value of the norm of $\mathbf{w}[n]$ becomes too small. Thus, we have

$$e_v^2 = \sum_{p=0}^{P} b_p^2 \sigma_v^2 = \sigma_c^2 \sigma_v^2 \tag{24}$$

where σ_c^2 is a constant. Now the cost function $J_2(\mathbf{w})$ becomes

$$J_2(\mathbf{w}) = \frac{E\{e^2[n]\} - \sigma_c^2 \sigma_v^2}{E\{y^2[n]\} - \sigma_v^2}. \tag{25}$$

Applying the standard gradient descent method to $J_2(\mathbf{w})$

$$\nabla_{\mathbf{w}} J_2 = \frac{2}{E\{y^2[n]\} - \sigma_v^2} \left(E\{e[n]\hat{\mathbf{x}}[n]\} \right.$$
$$\left. - \frac{E\{e^2[n]\} - \sigma_c^2 \sigma_v^2}{E\{y^2[n]\} - \sigma_v^2} E\{y[n]\mathbf{x}[n]\} \right) \tag{26}$$

where

$$\hat{\mathbf{x}}[n] = \mathbf{x}[n] - \sum_{p=1}^{P} b_p \mathbf{x}[n-p]. \tag{27}$$

The MSPE $E\{e^2[n]\}$ and the power of the output of the demixing vector $E\{y^2[n]\}$ can be estimated, respectively, by

$$\sigma_e^2[n] = \beta_e \sigma_e^2[n-1] + (1 - \beta_e)e^2[n]$$
$$\sigma_y^2[n] = \beta_y \sigma_y^2[n-1] + (1 - \beta_y)y^2[n] \tag{28}$$

where β_e and β_y are the corresponding forgetting factors with $0 \leq \beta_e, \beta_y < 1$.

Following standard stochastic approximation techniques [13], (26) yields the following online update:

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \frac{2\mu}{\sigma_y^2[n] - \sigma_v^2}$$
$$\times \left(e[n]\hat{\mathbf{x}}[n] - \frac{\sigma_e^2[n] - \sigma_c^2 \sigma_v^2}{\sigma_y^2[n] - \sigma_v^2} y[n]\mathbf{x}[n] \right) \tag{29}$$

where μ is the learning rate.

The above algorithm is a direct method, as there is no preprocessing involved. We next derive an associated algorithm based on prewhitening of the mixtures.

Notice that the correlation matrix $\mathbf{R}_{xx}[0]$ can be decomposed into the following form:

$$\mathbf{R}_{xx}[0] = \mathbf{E}\mathbf{D}_x \mathbf{E}^T \tag{30}$$

where $\mathbf{E}$ is an orthogonal matrix whose columns are the eigenvectors of $\mathbf{R}_{xx}[0]$, and $\mathbf{D}_x$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues. Multiplying the mixtures by the whitening matrix $\mathbf{P} = \mathbf{D}_x^{-(1/2)} \mathbf{E}^T$, we obtain the prewhitened vector $\tilde{\mathbf{x}}$

$$\tilde{\mathbf{x}} = \mathbf{D}_x^{-(1/2)} \mathbf{E}^T \mathbf{x}. \tag{31}$$

Then the new correlation matrix $\mathbf{R}_{\tilde{x}\tilde{x}}[0]$ is

$$\mathbf{P}_{\tilde{x}\tilde{x}}[0] = \mathbf{P}\mathbf{R}_{xx}[0]\mathbf{P}^T$$
$$= \mathbf{P}\mathbf{A}\mathbf{R}_s s[0]\mathbf{A}^T \mathbf{P}^T + \mathbf{P}\mathbf{R}_v \mathbf{P}^T$$
$$= \hat{\mathbf{A}}\mathbf{R}_{ss}[0]\hat{\mathbf{A}}^T + \hat{\mathbf{R}}_v = \mathbf{I} \tag{32}$$

with $\hat{\mathbf{A}} = \mathbf{P}\mathbf{A}$ and $\hat{\mathbf{R}}_v = \mathbf{P}\mathbf{R}_v \mathbf{P}^T$.

By virtue of prewhitening, the MPSE and the output power take the forms

$$E\{e^2[n]\} = e_s^2 + e_v^2$$
$$= \mathbf{w}^T \hat{\mathbf{A}}\hat{\mathbf{R}}_{ss}\hat{\mathbf{A}}^T \mathbf{w} + \sum_{p=0}^{P} b_p^2 \mathbf{w}^T \hat{\mathbf{R}}_v \mathbf{w} \tag{33}$$

$$E\{y^2[n]\} = \mathbf{w}^T \mathbf{R}_{\hat{x}\hat{x}}[0]\mathbf{w} = \mathbf{w}^T \mathbf{w}$$
$$= \mathbf{w}^T \hat{\mathbf{A}}\mathbf{R}_{ss}[0]\hat{\mathbf{A}}^T \mathbf{w} + \mathbf{w}^T \hat{\mathbf{R}}_v \mathbf{w}. \tag{34}$$

From (16), considering the multiplication of $\mathbf{P}$ with the mixing matrix $\mathbf{A}$, the cost function $J_2(\mathbf{w})$ now becomes

$$
\begin{aligned}
J_2(\mathbf{w}) &= \frac{(\mathbf{w}^T\mathbf{P})\mathbf{A}\hat{\mathbf{R}}_{ss}\mathbf{A}^T(\mathbf{P}^T\mathbf{w})}{(\mathbf{w}^T\mathbf{P})\mathbf{A}\mathbf{R}_{ss}[0]\mathbf{A}^T(\mathbf{P}^T\mathbf{w})} \\
&= \frac{\mathbf{w}^T\hat{\mathbf{A}}\hat{\mathbf{R}}_{ss}\hat{\mathbf{A}}^T\mathbf{w}}{\mathbf{w}^T\hat{\mathbf{A}}\mathbf{R}_{ss}[0]\hat{\mathbf{A}}^T\mathbf{w}} \\
&= \frac{E\{e^2[n]\} - \sum_{p=0}^{P} b_p^2 \mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w}}{E\{y^2[n]\} - \mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w}} \\
&= \frac{E\{e^2[n]\} - \sigma_c^2\mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w}}{\mathbf{w}^T\mathbf{w} - \mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w}} \\
&= \frac{E\{e^2[n]\} - \sigma_c^2\mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w}}{\mathbf{w}^T(\mathbf{I} - \hat{\mathbf{R}}_v)\mathbf{w}}.
\end{aligned}
\tag{35}
$$

Now, if we normalize the denominator $\mathbf{w}^T(\mathbf{I} - \hat{\mathbf{R}}_v)\mathbf{w}$ at each update of the adaptive algorithm by

$$
\mathbf{w}[n+1] \leftarrow \mathbf{w}[n+1]/\sqrt{\mathbf{w}^T(\mathbf{I} - \hat{\mathbf{R}}_v)\mathbf{w}}
\tag{36}
$$

then we have

$$
E\{y^2[n]\} - \mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w} = \mathbf{w}^T(\mathbf{I} - \hat{\mathbf{R}}_v)\mathbf{w} = 1.
\tag{37}
$$

This way, the cost function is simplified to

$$
\begin{aligned}
J_2(\mathbf{w}) &= E\{e^2[n]\} - \sigma_c^2\mathbf{w}^T\hat{\mathbf{R}}_v\mathbf{w} \\
&= E\{e^2[n]\} - \sigma_c^2(E\{y^2[n]\} - 1).
\end{aligned}
\tag{38}
$$

In this case, for a fixed linear predictor, applying the standard gradient descent method to $J_2(\mathbf{w})$, we can derive the following online update rule:

$$
\mathbf{w}[n+1] = \mathbf{w}[n] - 2\mu\left(e[n]\hat{\tilde{\mathbf{x}}}[n] - \sigma_c^2 y[n]\tilde{\mathbf{x}}[n]\right).
\tag{39}
$$

where $\hat{\tilde{\mathbf{x}}}[n]$ is given by

$$
\hat{\tilde{\mathbf{x}}}[n] = \tilde{\mathbf{x}}[n] - \sum_{p=1}^{P} b_p \tilde{\mathbf{x}}[n-p].
\tag{40}
$$

This update equation is followed by the normalization operation given by (36).

For both of the algorithms given in (29) and (39), the optimal choice of stepsizes and bounds on them are difficult problems, and further research is needed in this direction. Here we will choose them empirically in our simulations.

III. SIMULATIONS

Fig. 2 shows three source signals, denoted by s_0, s_1, s_2, used in simulations. These can be found in the ICALAB toolbox [3]. The coefficients of the linear predictor were randomly generated with a length of $P = 3$, and given by

$$
\mathbf{b} = [0.5080 \quad 0.2593 \quad -0.1340].
\tag{41}
$$

This way, the normalized prediction errors of the three signals, were, respectively, $\{1.3190, 0.7149, 0.1373\}$. The 4×3 mixing matrix $\mathbf{A}$ was randomly generated and given by

$$
\mathbf{A} = \begin{bmatrix}
-0.4312 & 0.1656 & -0.1342 \\
-0.0616 & -0.1530 & -0.5481 \\
-0.8704 & 0.0310 & 0.1596 \\
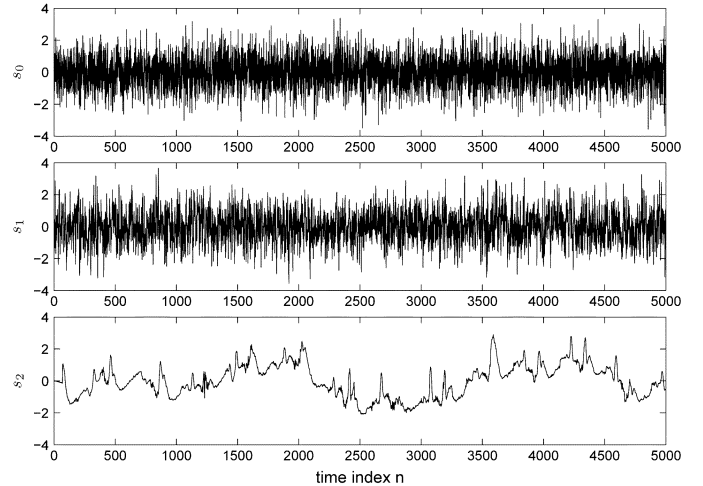0.9767 & -0.3321 & 0.5207
\end{bmatrix}.
\tag{42}
$$

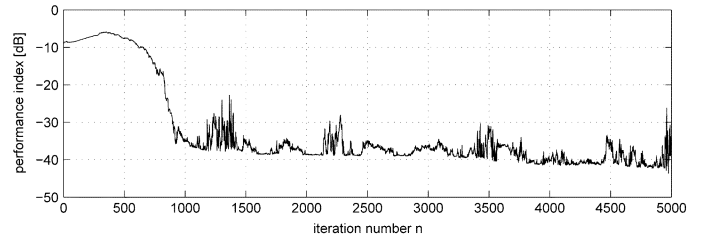Fig. 2. Tree source signals used in the simulations.

Fig. 3. Learning curve using the algorithm (29) with $\sigma_v^2 = 0.01$.

As we have one more mixture than the number of sources, this additional degree of freedom was used to estimate the variance of the additive white Gaussian noise. To illustrate the proposed approach, we used two different noise levels. For the first set of simulations, the variance of noise was $\sigma_v^2 = 0.01$. Each of source signals was normalized to a power of unity and the signal-to-noise ratio (SNR) in this sense was 20 dB. By minimizing the normalized MSPE, the signal with the smallest normalized prediction error will be extracted, which is the third signal s_2.

We first tested the direct algorithm given in (29). The forgetting factors were $\beta_e = \beta_y = 0.97$ and the stepsize $\mu = 0.001$. A learning curve for this case is shown in Fig. 3, with the performance index defined as [3]

$$
\text{PI} = 10\log_{10}\left(\frac{1}{M-1}\left(\sum_{m=0}^{M-1}\frac{g_m^2}{\max\{g_0^2, g_1^2, \ldots, g_{M-1}^2\}} - 1\right)\right)
\tag{43}
$$

with $\mathbf{g} = \mathbf{A}^T\mathbf{w} = [g_0 \ g_1 \ \cdots \ g_{M-1}]$. The resultant signal-to-interference-plus-noise ratio (SINR) was about 16.7 dB for the steady state and a learning curve of the resultant SINR for each iteration is shown by the solid line of Fig. 4.

As the performance index reached the level between -30 dB and -40 dB, we can say the signal s_2 had been extracted successfully. The associated prediction error during the adaptation represents a filtered version of the extracted source signal s_2 for the steady state.

To further illustrate the performance of the proposed algorithm, we next increased the variance of the noise to $\sigma_v^2 = 0.04$
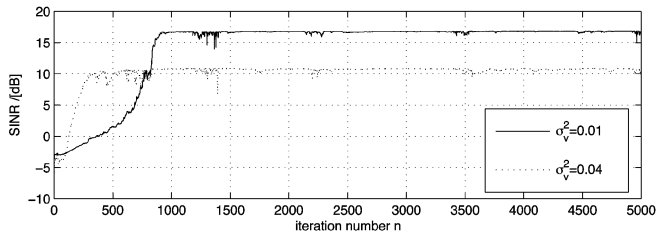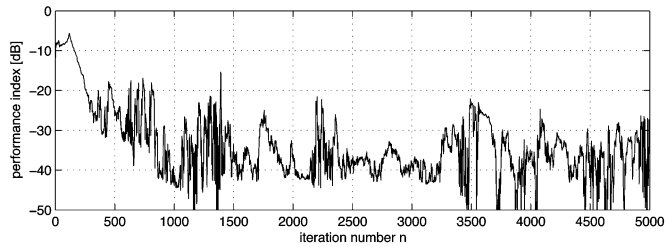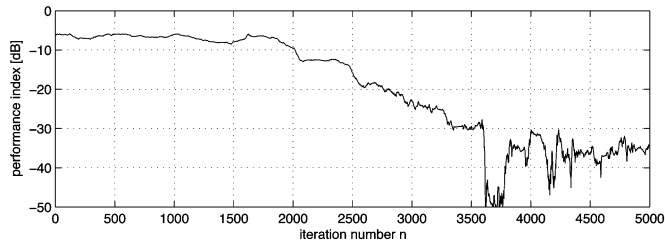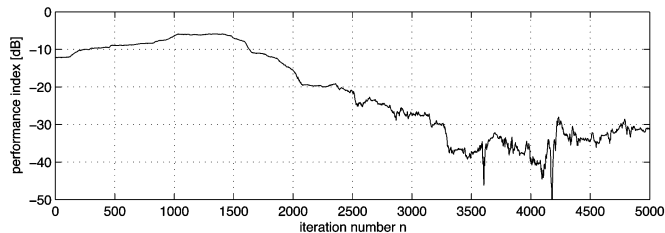
Fig. 4. Resultant SINR learning curves using algorithm (29).

Fig. 5. Learning curve using the algorithm (29) with $\sigma_v^2 = 0.04$.

Fig. 6. Learning curve using the algorithm (39) with $\sigma_v^2 = 0.01$.

Fig. 7. Learning curve using the algorithm (39) with $\sigma_v^2 = 0.04$.

and the resulting SNR was about 14 dB. The associated performance index learning curve is shown in Fig. 5 and the SINR learning curve in Fig. 4 by the dotted line. The resultant SINR value was approximately 10.7 dB. The value of the performance index at the steady state (below 30 dB) indicates a successful extraction.

Finally, we conducted the same experiments for the algorithm given in (39). The stepsize was chosen to be $\mu = 3.2$. It may seem surprising to have such a large stepsize, but this is a direct consequence of a very small instantaneous gradient value $2(e[n]\hat{x}[n] - \sigma_c y[n]\mathbf{x}[n])$, which is in the order of 10^{-3} during the adaptation. Moreover, we see the stepsize here is much larger (about 1000 times) than that in the direct approach for the first set of simulations. This is consistent with the observation for the noise-free BSE algorithms proposed in [12], where the stepsize with prewhitening is about 10^3 times the one with the normalized MPSE as the cost function. Thus, there is an close relationship between the role of prewhitening and the value of the
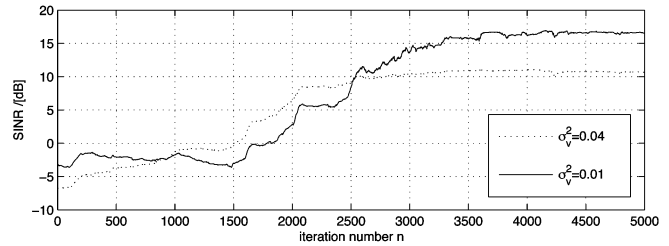

Fig. 8. Resultant SINR learning curves using the algorithm (39).

stepsize, which is the subject of further research. The performance index learning curves for the two different noise levels are shown, respectively, in Figs. 6 and 7. The corresponding SINR curves are shown in Fig. 8 with steady-state SINR values of about 16.6 and 10.7 dB. We can see in both cases the algorithm has extracted the source successfully.

IV. CONCLUSION

To extract the sources from their noisy measurements using a BSE architecture with a linear predictor, we have proposed a new cost function which accounts for the effect of noise, based on which two novel adaptive BSE algorithms have been derived. The first one is a direct algorithm and it works without prewhitening, but requires an estimation of some additional parameters during the adaptation. The second proposed algorithm requires a prewhitening operation, resulting in a simpler update. The proposed algorithms have been shown to operate effectively in noisy environments, as illustrated by simulations.

REFERENCES

[1] S. Haykin, Ed., *Unsupervised Adaptive Filtering, Volume 1: Blind Source Separation*. New York: Wiley, 2000.

[2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2003.

[4] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neur. Comput.*, vol. 9, pp. 1483–1492, 1997.

[5] A. Cichocki, R. Thawonmas, and S. Amari, "Sequential blind signal extraction in order specified by stochastics properties," *Electron. Lett.*, vol. 33, pp. 64–65, Jan. 1997.

[6] Y. Q. Li and J. Wang, "Sequential blind extraction of instantaneously mixed sources," *IEEE Trans. Signal Process.*, vol. 50, pp. 997–1006, May 2002.

[7] W. Liu and D. P. Mandic, "A normalized Kurtosis based algorithm for blind source extraction from noisey measurements," *Signal Process.*, vol. 86, no. 7, pp. 1580–1585, Jul. 2006.

[8] A. Cichocki and R. Thawonmas, "On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics," *Neural Process. Lett.*, vol. 12, pp. 91–98, 2000.

[9] A. K. Barros and A. Cichocki, "Extraction of specific signals with temporal structure," *Neur. Comput.*, vol. 13, pp. 1995–2003, 2001.

[10] A. Cichocki, T. Rutkowski, and K. Siwek, "Blind signal extraction of signals with specified frequency band," in *Proc. IEEE Workshop on Neural Netw. for Signal Process/*, Sep. 2002, pp. 515–524.

[11] D. P. Mandic and A. Cichocki, "An online algorithm for blind extraction of sources with different dynamical structures," in *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation*, Nara, Japan, Apr. 2003, pp. 645–650.

[12] W. Liu, D. P. Mandic, and A. Cichocki, "A class of novel blind source extraction algorithms based on a linear predictor," in *Proc. IEEE Int. Symp. Circuits and Systems*, Kobe, Japan, May 2005, pp. 3599–3602.

[13] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.